

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-297898  
(43)Date of publication of application : 12.11.1993

(51)Int.Cl. G10L 9/00  
G10L 9/18

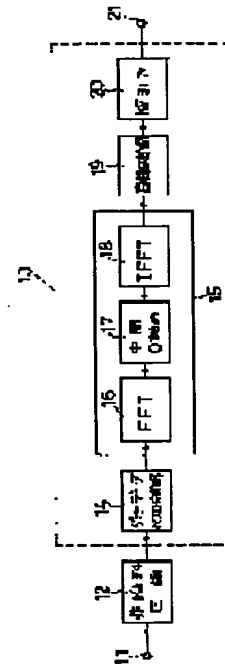
(21)Application number : 04-092263 (71)Applicant : SONY CORP  
(22)Date of filing : 18.03.1992 (72)Inventor : NISHIGUCHI MASAYUKI  
MATSUMOTO ATSUSHI

## (54) DATA QUANTITY CONVERTING METHOD

### (57)Abstract:

PURPOSE: To convert a variable number of input data into a fixed number of data without generating any ringing at a block end point in a nonlinear area.

CONSTITUTION: A nonlinear compression part 12 performs the nonlinear compression of a variable number of data by inputted blocks and a dummy data addition part 14 adds dummy data for interpolation from the final data value in the block to the starting data value to increase the number of data; and the band limiting type oversampling part 15 consisting of an FFT(fast Fourier transformation) processing part 16, an intermediate 0-filling process part 17, and an IFFT(inverse FFT) process part 18 performs oversampling, a linear interpolation part 19 performs linear interpolation, and a thinning-out process part 20 thins out the data to convert the data into a fixed number of sample data.



## LEGAL STATUS

[Date of request for examination] 20.01.1999  
[Date of sending the examiner's decision of rejection]  
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]  
[Date of final disposal for application]  
[Patent number]  
[Date of registration]  
[Number of appeal against examiner's decision of rejection]  
[Date of requesting appeal against examiner's decision of rejection]  
[Date of extinction of right]

(19)日本国特許庁 (J P)

## (12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平5-297898

(43)公開日 平成 5 年(1993)11月12日

(51)Int.Cl.<sup>5</sup>G 1 0 L 9/00  
9/18

識別記号

庁内整理番号

M 8946-5H  
E 8946-5H

F I

技術表示箇所

審査請求 未請求 請求項の数 2 (全 16 頁)

(21)出願番号 特願平4-92263

(22)出願日 平成 4 年(1992) 3 月18日

(71)出願人 000002185

ソニー株式会社

東京都品川区北品川 6 丁目 7 番35号

(72)発明者 西口 正之

東京都品川区北品川 6 丁目 7 番35号 ソニ  
ー株式会社内

(72)発明者 松本 淳

東京都品川区北品川 6 丁目 7 番35号 ソニ  
ー株式会社内

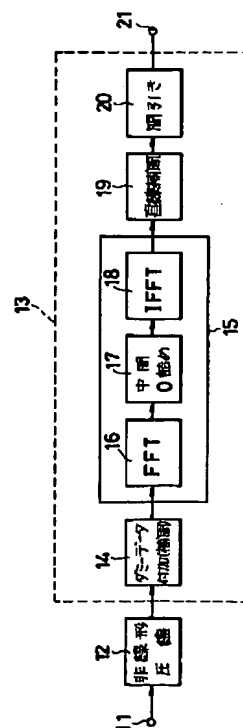
(74)代理人 弁理士 小池 晃 (外 2 名)

(54)【発明の名称】 データ数変換方法

(57)【要約】

【構成】 入力されたブロック毎に可変個数のデータを、非線形圧縮部 12 で非線形圧縮し、ダミーデータ付加部 14 でブロック内の最後のデータ値から最初のデータ値までの補間するようなダミーデータを付加してデータ個数を拡大し、FFT (高速フーリエ変換) 処理部 16、中間 0 詰め処理部 17、IFFT (逆FFT) 処理部 18 から成る帯域制限型オーバーサンプリング部 15 でオーバーサンプリングし、直線補間 19 で直線補間し、間引き処理部 20 で間引くことにより一定個数のサンプルデータに変換する。

【効果】 可変個数の入力データを、非線形領域で、ブロック端点にリングング等が発生しないように、一定の個数のデータに変換できる。



## 【特許請求の範囲】

【請求項1】 ブロック内の波形データ又は波形を表すパラメータデータの個数が可変とされたデータを非線形圧縮する工程と、

このブロック毎に可変の個数の非線形圧縮データをブロック毎に一定の個数の基準データとを比較するために上記可変個数の非線形圧縮データを上記一定個数に変換する工程とを有することを特徴とするデータ数変換方法。

【請求項2】 上記ブロック毎に可変の個数の非線形圧縮データに、ブロック内の最後のデータからブロック内の最初のデータまでの値を補間するようなダミーデータを付加してデータ個数を拡大した後、帯域制限型のオーバーサンプリングを施すことを特徴とする請求項1記載の高効率符号化方法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、データ数変換方法に関し、特に、音声合成分析装置（ボコーダ）等において算出されたスペクトルの振幅データのような可変個数のデータを一定個数のデータに変換するようなデータ数変換方法に関する。

## 【0002】

【従来の技術】 オーディオ信号（音声信号や音響信号を含む）の時間領域や周波数領域における統計的性質と人間の聴感上の特性を利用して信号圧縮を行うような符号化方法が種々知られている。この符号化方法としては、大別して時間領域での符号化、周波数領域での符号化、分析合成符号化等が挙げられる。

【0003】 音声信号等の高効率符号化の例として、MBE（Multiband Excitation: マルチバンド励起）符号化、SBE（Singleband Excitation: シングルバンド励起）符号化、ハーモニック（Harmonic）符号化、SBC（Sub-band Coding: 帯域分割符号化）、LPC（Linear Predictive Coding: 線形予測符号化）、あるいはDCT（離散コサイン変換）、MDCT（モデファイドDCT）、FFT（高速フーリエ変換）等において、スペクトル振幅やそのパラメータ（LSPパラメータ、 $\alpha$ パラメータ、 $k$ パラメータ等）のような各種情報データを量子化する場合に、従来においてはスカラ量子化を行うことが多い。

## 【0004】

【発明が解決しようとする課題】 ところで、ビットレートを例えば3～4 kbps 程度にまで低減し、量子化効率を更に向上させようとする、スカラ量子化では量子化雑音（歪み）が大きくなってしまい、実用化が困難であった。そこで、これらの符号化の際に得られる時間軸データや周波数軸データやフィルタ係数データ等を個々に量子化せず、複数のデータを組（ベクトル）にまとめて一つの符号で表現して量子化するベクトル量子化が注目されている。

【0005】 しかしながら、上記MBE、SBE、LPC等のスペクトル振幅データ等は、ピッチに依存して個数が増えるため、そのままベクトル量子化しようとすると可変次元のベクトル量子化が必要となり、構成が複雑化するのみならず、良好な特性を得ることが困難である。

【0006】 また、量子化の前にデータのブロック（フレーム）間差分をとるような場合にも、前後のブロック（フレーム）内のデータの個数が一致していないと、差分をとることができない。このように、可変個数のデータを一定個数に変換することがデータ処理の過程で必要とされることがあるが、特性の良好なデータ数変換が望まれる。

【0007】 本発明は、このような実情に鑑みてなされたものであり、可変個数のデータを一定個数に変換することができ、端点でリングング等の発生しない特性の良好なデータ数変換が行えるようなデータ数変換方法の提供を目的とする。

## 【0008】

【課題を解決するための手段】 本発明に係るデータ数変換方法は、ブロック内の波形データ又は波形を表すパラメータデータの個数が可変とされたデータを非線形圧縮する工程と、このブロック毎に可変の個数の非線形圧縮データをブロック毎に一定の個数の基準データと非線形領域で比較するために上記可変個数の非線形圧縮データを上記一定個数に変換するデータ数変換器を用いる工程とを有することにより、上述の課題を解決する。

【0009】 ここで、上記ブロック毎に可変の個数の非線形圧縮データに、ブロック内の最後のデータからブロック内の最初のデータまでの値を補間するようなダミーデータを付加してデータ個数を拡大した後、帯域制限型のオーバーサンプリングを施すようにするのが好ましい。このブロック内最後のデータからブロック内最初のデータまでの値を補間するようなダミーデータとは、ブロックの端点に急激な値の変化をもたらさないような、あるいは値が跳んだり不連続となったりしないようなデータのことであり、ブロック内最後のデータ値を一定区間保持した後、ブロック内最初のデータ値に変化させ、このブロック内最初のデータ値を一定区間保持するような値の変化の形態が挙げられる。上記帯域制限型のオーバーサンプリングは、FFT（高速フーリエ変換）等の直交変換をし、オーバーサンプリングの倍数に応じた区間だけ0を詰め（あるいはローパスフィルタ処理を施し）た後、IFFT（逆FFT）等の逆直交変換をすればよい。

【0010】 上記非線形圧縮されるデータとしては、音声信号や音響信号等のオーディオ信号を、周波数軸上データに変換したものをを用いることができ、具体的には、例えば、MBE（Multiband Excitation: マルチバンド励起）符号化の場合のスペクトル包絡振幅データや、S

BE (Singleband Excitation: シングルバンド励起) 符号化、ハーモニック (Harmonic) 符号化、SBC (Sub-band Coding: 帯域分割符号化)、LPC (Linear Predictive Coding: 線形予測符号化)、あるいはDCT (離散コサイン変換)、MDCT (モデファイドDCT)、FFT (高速フーリエ変換) 等におけるスペクトル振幅データやそのパラメータ (LSPパラメータ、 $\alpha$ パラメータ、 $k$ パラメータ等) データ等を用いることができる。また、一定個数に変換されたデータを、ベクトル量子化することが考えられ、このベクトル量子化の前に、ブロック毎に一定個数のデータのブロック間の差分をとり、このブロック間差分データに対してベクトル量子化を施すようにしてもよい。

#### 【0011】

【作用】一定個数の非線形圧縮データに変換されて基準データと非線形領域で比較することができ、ブロック間差分をとってベクトル量子化することが可能となる。また、データ数変換前のブロック内のデータ値の連続性を高め、ブロック端点でリングング等の生じない品質の高いデータ数変換が行える。

#### 【0012】

【実施例】以下、本発明に係るデータ数変換方法の実施例について、図面を参照しながら説明する。図1は本発明の一実施例となるデータ数変換方法の概略構成を示している。

【0013】図1において、入力端子11には、後述するMBEボコーダにより算出されたスペクトルエンベロープの振幅データ等が供給されている。この振幅データは、例えば図2のAに示すようなスペクトルを有する音声信号を分析して、ピッチ周波数 (角周波数)  $\omega$  を求め、このピッチ周波数  $\omega$  に応じたスペクトルの周期性を

$$a_{dB}(m) = 20 \log_{10} a(m)$$

である。この対数変換された振幅データ  $a_{dB}(m)$  の個数  $m_{MX}+1$  は、上述したようにピッチに依存して変化するため、一定個数 ( $M$ 個) の振幅データ  $b_{dB}(m)$  に変換する。これは一種のサンプリングレート (サンプルレート) 変換である。なお、非線形圧縮部12での圧縮処理は、dB領域への対数圧縮の他に、例えばいわゆる  $\mu$ -law や  $\alpha$ -law のような疑似対数圧縮処理を施してもよい。このように、振幅を圧縮することにより、能率的な符号化が実現される。

【0016】MBEボコーダに入力される時間軸上の音声信号に対するサンプリング周波数  $f_s$  は、通常8 kHzで、全帯域幅は3.4 kHz (ただし有効帯域は200~3400 Hz) であり、女声の高い方から男声の低い方までのピッチラグ (ピッチ周期に相当するサンプル数) は、20~147程度である。従って、ピッチ (角) 周

考慮して、各高調波 (ハーモニックス) 位置での振幅を求め、図2のBに示すようなスペクトル包絡 (エンベロープ) を表す振幅データを求める場合に、一定の有効帯域 (例えば200~3400 Hz) 内でのこの振幅データの個数は、上記ピッチ周波数  $\omega$  に依存して変化する。そこで、一定の固定周波数 (角周波数)  $\omega_c$  を想定し、この一定周波数  $\omega_c$  の各高調波位置での上記スペクトル包絡の振幅データを求めることで、データ個数を一定とするものである。

【0014】図1の例では、入力端子11からの可変個数 ( $m_{MX}+1$  個) の入力データを、非線形圧縮部12にて例えばdB領域に圧縮 (対数圧縮) した後、データ個数変換本体部13にて一定個数 ( $M$ 個) のデータに変換している。データ個数変換本体部13は、ダミーデータ付加部14と帯域制限型オーバーサンプリング部15とを有し、帯域制限型オーバーサンプリング部15は直交変換、例えばFFT (高速フーリエ変換) 処理部16、中間0詰め処理部17、及び逆直交変換、例えばIFFT (逆FFT) 処理部18から成っている。帯域制限型オーバーサンプリングが施されたデータは、直線補間部19で直線補間され、間引き処理部20で間引かれて、一定個数のデータとなり、出力端子21から取り出される。

【0015】ここで、後述するMBEボコーダにおいて算出される  $m_{MX}+1$  個の振幅データ列を  $a(m)$  とする。 $m$  は上記高調波 (ハーモニックス) の次数あるいはバンド番号であり、 $m_{MX}$  が最大値であるが、 $m=0$  のバンドの振幅データも含めて、全バンドの振幅データの個数は  $m_{MX}+1$  個となる。この振幅データ  $a(m)$  を、非線形圧縮部14にて例えばdB領域に変換する。すなわち得られたデータを  $a_{dB}(m)$  とするとき、

$$\dots (1)$$

波数  $\omega$  は、 $8000/147 \approx 54$  (Hz) から  $8000/20 = 400$  (Hz) 程度までの間で変動することになる。従って、周波数軸上で上記3.4 kHzまでの間に約8~63本のピッチパルス (ハーモニックス) が立つことになる。すなわち、周波数軸上のdB領域の波形として、8サンプル乃至63サンプルから成るデータを、一定のサンプル数、例えば44サンプルに、サンプル数変換を行うわけである。これが、図2のCに示すように、一定のピッチ周波数 (角周波数)  $\omega_c$  毎のハーモニックスの位置のサンプルを求めることに相当する。

【0017】次にダミーデータ付加部14において、上記  $m_{MX}+1$  個の圧縮データ  $a_{dB}(m)$  をFFTし易い数 ( $2$  のべき乗等)  $N_F$  (例えば  $N_F = 256$ ) に延長する。すなわち、 $m_{MX}+1$  から  $N_F$  までの区間のダミーデータ  $a'_{dB}(m)$  として、例えば

$$m_{MX}+1 \leq m < N_F / 2 : a'_{dB}(m) = a_{dB}(m_{MX})$$

$$N_F / 2 \leq m < 3 N_F / 4 :$$

$$a'_{dB}(m) = a_{dB}(m_{MX}) \times k_1 + a_{dB}(0) \times k_2$$

$$\begin{aligned} \text{ただし、} k_1 &= (3N_F / 4 - n) / (N_F / 4) \\ k_2 &= (n - N_F / 2) / (N_F / 4) \\ 3N_F / 4 \leq m < N_F : a'_{dB}(m) &= a_{dB}(0) \end{aligned}$$

... (2)

を用いて延長する。これは、図3に示すように、 $0 \sim m_{MX}$ の区間に元の振幅データ  $a_{dB}(m)$  を配置し、 $m_{MX} + 1 \leq m < N_F / 2$  の区間ではブロック内の最後のデータである  $a_{dB}(m_{MX})$  を保持し、 $N_F / 2 \leq m < 3N_F / 4$  の区間は直線的に補間し、 $3N_F / 4 \leq m < N_F$  の区間ではブロック内の最初のデータ  $a_{dB}(0)$  を保持するような折線となる。

【0018】すなわち、図3に示すレート変換する原波形の左端と右端が徐々につながるようにデータを作り込んで詰めてゆくのである。FFTでは、変換前の波形を図3の破線に示すように繰り返し波形とみなしているの、 $m = N_F$  の点は、 $m = 0$  に接続されることになる。

【0019】これは、FFT後に周波数軸上で乗算を行うようなフィルタリングを行うと、元の図3に示す軸上でたたみ込みが行われることになる。このため、図4に示すように原波形以外の部分 ( $m_{MX} < m < N_F$  の区間) に単純に0詰めを行うと、不連続点にて図4の破線Rに示すようなリングングが発生し、良好なレート変換が行われなくなる。このような不具合を防止するために、上記図3に示すように、ブロック端点に急激な値の変化をもたらさないようにダミーデータを詰めてゆくわけである。なお、上記ダミーデータの具体例の他に、図3の破線Iに示すようにブロックの最後のデータからブロックの最初のデータまでの全体を直線的に補間するようになり、曲線的に補間するようにしてもよい。

【0020】次に、上記  $N_F$  点 ( $N_F$  サンプル) に拡張された数列 (データ列) に対して、帯域制限型オーバーサンプリング部15のFFT (高速フーリエ変換) 処理部16により、 $N_F$  点FFTを施し、図5のAに示すような  $0 \sim N_F$  の数列 (スペクトル) を得る。この数列の  $0 \sim \pi$  に相当する部分と、 $\pi \sim 2\pi$  に相当する部分との間に、中間0詰め処理部17により、 $(O_S - 1) N_F$  個の0を詰める。このときの  $O_S$  が、オーバーサンプリングの比率 (レシオ) になる。例えば  $O_S = 8$  の場合には、図5のBに示すように上記数列の  $0 \sim \pi$  相当部分と  $\pi \sim 2\pi$  相当部分との間に  $7N_F$  個の0を詰め、 $8N_F$  点 (例えば  $N_F = 256$  のとき  $2048$  点) の数列とする。

【0021】ここで、上記0詰め操作は、LPF処理とみることでもできる。すなわち、サンプリングレートとして  $O_S N_F$  の数列に対し、 $O_S N_F$  で動作しているデジタルフィルタが、図6のAの太線で示すような  $\pi/8$  のカットオフでローパス処理して、図6のBに示すようなサンプル列を得ることになる。このフィルタ操作において、上記図4の破線Rで示したようなリングングが発生する虞れがある。本実施例では、そのリングングの発

生を防ぐために、原波形の左端と右端とをなだらかに (微分係数の急激な変化がないように) つなぐようにしている。

【0022】次に、IFFT (逆FFT) 処理部18により、 $O_S N_F$  点 (例えば  $2048$  点) の逆FFTを行うと、 $O_S$  倍にオーバーサンプルされた図7に示すような振幅データ (ダミーデータも含む) が得られる。このデータ列の有効な部分、すなわち  $0 \sim O_S \times (m_{MX} + 1)$  を取り出すと、原波形 (元の振幅データ  $a_{dB}(m)$ ) が  $O_S$  倍の密度となってオーバーサンプルされたものが得られる。これはまだピッチに応じて可変の個数  $m_{MX} + 1$  に依存するデータ列である。

【0023】次に、これを固定のデータ数に変換するために、直線補間を行う。例えば、図8のAは、上記  $m_{MX} = 19$  (変換前の全バンド数、振幅データの個数が  $20$ ) の場合を示しており、これを、 $O_S = 8$  として8倍オーバーサンプリングすることにより、図8のBに示すように、 $0 \sim \pi$  の間に  $O_S \times (m_{MX} + 1) = 160$  個のサンプルデータが存在するわけであるが、これを直線補間部19で一定個数  $N_M$ 、例えば  $N_M = 2048$  に直線補間する。

【0024】図9のAは、直線補間部19により直線補間されて得られた一定個数  $N_M$  (例えば  $N_M = 2048$ ) のデータを示しており、この  $2048$  サンプルのデータを所定のサンプル数  $M$  (例えば  $M = 44$ ) に変換するために、間引き処理部20により  $2048$  点のデータを間引いて  $44$  点のデータを得ている。ここで、第  $0 \sim$  第  $2047$  番目のサンプルデータの内、DC値 (直流データ値、第0番目のデータ値) は伝送する必要がないので、 $\text{rint} (2048 / 44) \cdot i$  の値を間引き値として採用し、 $44$  個のデータを得るようにすればよい。ただし、 $1 \leq i \leq 44$  であり、 $\text{rint}$  は、nearest integer、すなわち最も近い整数値を示す関数である。

【0025】このようにして、一定サンプル数  $M$  個のデータに変換した数列  $b_{dB}(n)$ 、ただし  $1 \leq n \leq M$ 、を得る。この固定データ数の数列を、必要に応じてブロック間、あるいはフレーム間で差分をとり、ベクトル量子化を施して、そのインデックスを伝送するようになればよい。

【0026】受信側 (合成側あるいはデコーダ側) では、そのインデックスより、ベクトル量子化及び逆量子化された数列  $b_{vQdB}(n)$  なる  $M$  点の波形データを得る。そのデータ列を、同様の方法で、すなわち帯域制限オーバーサンプリング、直線補間、間引きのそれぞれ逆の操作を施すことにより、必要なポイント数の上記  $m_{MX} + 1$

点の数列に変換する。なお、 $m_{MX}$  (あるいは $m_{MX}+1$ ) は、別途伝送するピッチ情報により求めることができる。例えばサンプリング周期に対して規格化されたピッチ周期を $p$ とすると、ピッチ周波数(角周波数) $\omega$ は、 $2\pi/p$ で求められ、 $\pi/\omega=p/2$ より、 $m_{MX}+1=\text{inint}(p/2)$ として算出できる。この $m_{MX}+1$ 点の振幅情報をもとにしてデコード処理を行う。

【0027】次に、上述したようなデータ数変換方法が適用可能な、音声信号の合成分析符号化装置(いわゆるボコーダ)の一種のMBE(Multiband Excitation: マルチバンド励起)ボコーダの具体例について、図面を参照しながら説明する。以下に説明するMBEボコーダは、D.W. Griffin and J.S. Lim, "Multiband Excitation Vocoder," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, No. 8, pp. 1223-1235, Aug. 1988に開示されているものであり、従来のPARCOR(Partial auto-CORrelation: 偏自己相関)ボコーダ等では、音声のモデル化の際に有声音区間と無声音区間とをブロックあるいはフレーム毎に切り換えていたのに対し、MBEボコーダでは、同時刻(同じブロックあるいはフレーム内)の周波数軸領域に有声音(Voiced)区間と無声音(Unvoiced)区間とが存在するという仮定でモデル化している。

【0028】図10は、上記MBEボコーダに本発明を

$$x_w(k, q) = x(q) w(kL - q)$$

となる。この(1)式において、 $k$ はブロック番号を、 $q$ はデータの時間インデックス(サンプル番号)を表し、処理前の入力信号の $q$ 番目のデータ $x(q)$ に対して第 $k$ ブロックの窓(ウィンドウ)関数 $w(kL - q)$ により窓

$$w_r(r) = \begin{cases} 1 & 0 \leq r < N \\ 0 & r < 0, N \leq r \end{cases} \quad \dots (4)$$

また、窓かけ処理部104での図11のBに示すような

$$w_h(r) = \begin{cases} 0.54 - 0.46 \cos(2\pi r/(N-1)) & 0 \leq r < N \\ 0 & r < 0, N \leq r \end{cases}$$

である。このような窓関数 $w_r(r)$ あるいは $w_h(r)$ を用いるときの上記(3)式の窓関数 $w(r)$ (= $w(kL - q)$ )の非零区間は、

$$0 \leq kL - q < N$$

これを變形して、

$$kL - N < q \leq kL$$

従って例えば上記方形窓の場合に窓関数 $w_r(kL - q) = 1$ となるのは、図12に示すように、 $kL - N < q \leq kL$ のときとなる。また、上記(3)~(5)式は、長さ $N$ (=256)サンプルの窓が、 $L$ (=160)サンプルずつ前進してゆくことを示している。以下、上記(4)式、(5)式の各窓関数で切り出された各 $N$ 点( $0 \leq r < N$ )の非零サンプル列を、それぞれ $x_{wr}(k, r)$ 、 $x_{wh}(k, r)$ と表すことにする。

【0030】窓かけ処理部104では、図13に示すよ

適用した実施例の全体の概略構成を示すブロック図である。この図10において、入力端子101には音声信号が供給されるようになっており、この入力音声信号は、HPF(ハイパスフィルタ)等のフィルタ102に送られて、いわゆるDC(直流)オフセット分の除去や帯域制限(例えば200~3400Hzに制限)のための少なくとも低域成分(200Hz以下)の除去が行われる。このフィルタ102を介して得られた信号は、ピッチ抽出部103及び窓かけ処理部104にそれぞれ送られる。ピッチ抽出部103では、入力音声信号データが所定サンプル数 $N$ (例えば $N=256$ )単位でブロック分割され(あるいは方形窓による切り出しが行われ)、このブロック内の音声信号についてのピッチ抽出が行われる。このような切り出しブロック(256サンプル)を、例えば図11のAに示すように $L$ サンプル(例えば $L=160$ )のフレーム間隔で時間軸方向に移動させており、各ブロック間のオーバーラップは $N-L$ サンプル(例えば96サンプル)となっている。また、窓かけ処理部104では、1ブロック $N$ サンプルに対して所定の窓関数、例えばハミング窓をかけ、この窓かけブロックを1フレーム $L$ サンプルの間隔で時間軸方向に順次移動させている。

【0029】このような窓かけ処理を数式で表すと、

$$\dots (3)$$

かけ処理されることによりデータ $x_w(k, q)$ が得られることを示している。ピッチ抽出部103内での図11のAに示すような方形窓の場合の窓関数 $w_r(r)$ は、

ハミング窓の場合の窓関数 $w_h(r)$ は、

$$\begin{cases} 0 \leq r < N \\ r < 0, N \leq r \end{cases} \quad \dots (5)$$

うに、上記(5)式のハミング窓がかけられた1ブロック256サンプルのサンプル列 $x_{wh}(k, r)$ に対して1792サンプル分の0データが付加されて(いわゆる0詰めされて)2048サンプルとされ、この2048サンプルの時間軸データ列に対して、直交変換部105により例えばFFT(高速フーリエ変換)等の直交変換処理が施される。

【0031】ピッチ抽出部103では、上記 $x_{wr}(k, r)$ のサンプル列(1ブロック $N$ サンプル)に基づいてピッチ抽出が行われる。このピッチ抽出法には、時間波形の周期性や、スペクトルの周期的周波数構造や、自己相関関数を用いるもの等が知られているが、本実施例では、センタクリップ波形の自己相関法を採用している。このときのブロック内でのセンタクリップレベルについては、1ブロックにつき1つのクリップレベルを設定して

もよいが、ブロックを細分割した各部（各サブブロック）の信号のピークレベル等を検出し、これらの各サブブロックのピークレベル等の差が大きいときに、ブロック内でクリップレベルを段階的にあるいは連続的に変化させるようにしている。このセンタクリップ波形の自己相関データのピーク位置に基づいてピーク周期を決めている。このとき、現在フレームに属する自己相関データ

（自己相関は1ブロックNサンプルのデータを対象として求められる）から複数のピークを求めておき、これらの複数のピークの内の最大ピークが所定の閾値以上のときには該最大ピーク位置をピッチ周期とし、それ以外のときには、現在フレーム以外のフレーム、例えば前後のフレームで求められたピッチに対して所定の関係を満たすピッチ範囲内、例えば前フレームのピッチを中心として $\pm 20\%$ の範囲内にあるピークを求め、このピーク位置に基づいて現在フレームのピッチを決定するようにしている。このピッチ抽出部103ではオープンループによる比較的ラフなピッチのサーチが行われ、抽出されたピッチデータは高精度（ファイン）ピッチサーチ部10

$$S(j) = H(j) | E(j) |$$

と表現するようなモデルを想定している。ここで、 $J$  は  $\pi \omega_s = f_s / 2$  に対応し、サンプリング周波数  $f_s = 2 \pi \omega_s$  が例えば8 kHzのときには4 kHzに対応する。上記(6)式中において、周波数軸上のスペクトルデータ  $S(j)$  が図14のAに示すような波形のとき、 $H(j)$  は、図14のBに示すような元のスペクトルデータ  $S(j)$  のスペクトル包絡線（エンベロープ）を示し、 $E(j)$  は、図14のCに示すような等レベルで周期的な励起信号（エキサイテーション）のスペクトルを示している。すなわち、FFTスペクトル  $S(j)$  は、スペクトルエンベロープ  $H(j)$  と励起信号のパワースペクトル  $|E(j)|$  との積としてモデル化される。

【0034】上記励起信号のパワースペクトル  $|E(j)|$  は、上記ピッチに応じて決定される周波数軸上の波形の周期性（ピッチ構造）を考慮して、1つの帯域（バンド）の波形に相当するスペクトル波形を周波数軸上の各

6に送られて、クローズドループによる高精度のピッチサーチ（ピッチのファインサーチ）が行われる。

【0032】高精度（ファイン）ピッチサーチ部106には、ピッチ抽出部103で抽出された整数（インテジャ）値の粗（ラフ）ピッチデータと、直交変換部105により例えばFFTされた周波数軸上のデータとが供給されている。この高精度ピッチサーチ部106では、上記粗ピッチデータ値を中心に、0.2~0.5きざみで±数サンプルずつ振って、最適な小数点付き（フローティング）のファインピッチデータの値へ追い込む。このときのファインサーチの手法として、いわゆる合成による分析（Analysis by Synthesis）法を用い、合成されたパワースペクトルが原音のパワースペクトルに最も近くなるようにピッチを選んでいく。

【0033】このピッチのファインサーチについて説明する。まず、上記MBEポコードにおいては、上記FFT等により直交変換された周波数軸上のスペクトルデータとしての  $S(j)$  を

$$0 < j < J \quad \dots (6)$$

バンド毎に繰り返すように配列することにより形成される。この1バンド分の波形は、例えば上記図13に示すような256サンプルのハミング窓関数に1792サンプル分の0データを付加（0詰め）した波形を時間軸信号と見なしてFFTし、得られた周波数軸上のある帯域幅を持つインパルス波形を上記ピッチに応じて切り出すことにより形成することができる。

【0035】次に、上記ピッチに応じて分割された各バンド毎に、上記  $H(j)$  を代表させるような（各バンド毎のエラーを最小化するような）値（一種の振幅） $|A_m|$  を求める。ここで、例えば第  $m$  バンド（第  $m$  高調波の帯域）の下限、上限の点をそれぞれ  $a_m$ 、 $b_m$  とするとき、この第  $m$  バンドのエラー  $\varepsilon_m$  は、

【0036】

【数1】

$$\varepsilon_m = \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} \quad \dots (7)$$

で表せる。このエラー  $\varepsilon_m$  を最小化するような  $|A_m|$  は、

【0037】

【数2】

$$\frac{\partial \varepsilon_m}{\partial |A_m|} = 2 \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} |E(j)|$$

$$\therefore |A_m| = 2 \sum_{j=a_m}^{b_m} |S(j)| |E(j)| / \sum_{j=a_m}^{b_m} |E(j)|^2 \quad \dots (8)$$

となり、この(8)式の  $|A_m|$  のとき、エラー  $\varepsilon_m$  を最小化する。このような振幅  $|A_m|$  を各バンド毎に求

め、得られた各振幅  $|A_m|$  を用いて上記(7)式で定義された各バンド毎のエラー  $\varepsilon_m$  を求める。次に、この

ような各バンド毎のエラー  $\varepsilon_m$  の全バンドの総和値  $\Sigma \varepsilon_m$  を求める。さらに、このような全バンドのエラー総和値  $\Sigma \varepsilon_m$  を、いくつかの微小に異なるピッチについて求め、エラー総和値  $\Sigma \varepsilon_m$  が最小となるようなピッチを求める。

【0038】すなわち、上記ピッチ抽出部103で求められたラフピッチを中心として、例えば0.25きざみで上下に数種類ずつ用意する。これらの複数種類の微小に異なるピッチの各ピッチに対してそれぞれ上記エラー総和値  $\Sigma \varepsilon_m$  を求める。この場合、ピッチが定まるとバンド幅が決まり、上記(8)式より、周波数軸上データのパワースペクトル  $|S(j)|$  と励起信号スペクトル  $|E(j)|$  とを用いて上記(7)式のエラー  $\varepsilon_m$  を求め、その全バンドの総和値  $\Sigma \varepsilon_m$  を求めることができる。このエラー総和値  $\Sigma \varepsilon_m$  を各ピッチ毎に求め、最小となるエラー総和値に対応するピッチを最適のピッチとして決定するわけである。以上のようにして高精度ピッチサーチ部106で最適のファイン(例えば0.25きざみ)ピッ

$$NSR = \frac{\sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \}^2}{\sum_{j=a_m}^{b_m} |S(j)|^2} \dots (9)$$

と表せ、このNSR値が所定の閾値(例えば0.3)より大のとき(エラーが大きき)ときには、そのバンドでの  $|A_m| |E(j)|$  による  $|S(j)|$  の近似が良くない(上記励起信号  $|E(j)|$  が基底として不適当である)と判断でき、当該バンドをUV(Unvoiced、無声音)と判別する。これ以外のときは、近似がある程度良好に行われていると判断でき、そのバンドをV(Voiced、有声音)と判別する。

【0042】次に、振幅再評価部108には、直交変換部105からの周波数軸上データ、高精度ピッチサーチ

チが求められ、この最適ピッチに対応する振幅  $|A_m|$  が決定される。

【0039】以上ピッチのファインサーチの説明においては、説明を簡略化するために、全バンドが有声音(Voiced)の場合を想定しているが、上述したようにMBEボコーダにおいては、同時刻の周波数軸上に無声音(Unvoiced)領域が存在するというモデルを採用していることから、上記各バンド毎に有声音/無声音の判別を行うことが必要とされる。

【0040】上記高精度ピッチサーチ部106からの最適ピッチ及び振幅  $|A_m|$  のデータは、有声音/無声音判別部107に送られ、上記各バンド毎に有声音/無声音の判別が行われる。この判別のために、NSR(ノイズtoシグナル比)を利用する。すなわち、第mバンドのNSRは、

【0041】

【数3】

部106からのファインピッチと評価された振幅  $|A_m|$  との各データ、及び上記有声音/無声音判別部107からのV/UV(有声音/無声音)判別データが供給されている。この振幅再評価部108では、有声音/無声音判別部107において無声音(UV)と判別されたバンドに関して、再度振幅を求めている。このUVのバンドについての振幅  $|A_m|_{UV}$  は、

【0043】

【数4】

$$|A_m|_{UV} = \sqrt{\sum_{j=a_m}^{b_m} |S(j)|^2 / (b_m - a_m + 1)} \dots (10)$$

にて求められる。

【0044】この振幅再評価部108からのデータは、データ数変換(一種のサンプリングレート変換)部109に送られる。このデータ数変換部109は、上記ピッチに応じて周波数軸上での分割帯域数が異なり、データ数(特に振幅データの数)が異なることを考慮して、一定の個数にするためのものである。すなわち、例えば有効帯域を3400kHzまでとすると、この有効帯域が上記ピッチに応じて、8バンド~63バンドに分割されることになり、これらの各バンド毎に得られる上記振幅  $|A_m|$  (UVバンドの振幅  $|A_m|_{UV}$  も含む)データの

個数  $m_{MX} + 1$  も8~63と変化することになる。このためデータ数変換部109では、この可変個数  $m_{MX} + 1$  の振幅データを一定個数M(例えば44個)のデータに変換している。

【0045】ここで本実施例においては、上記図1~図9と共に説明したように、周波数軸上の有効帯域1ブロック分の振幅データに対して、ブロック内の最後のデータからブロック内の最初のデータまでの値を補間するようなダミーデータを付加してデータ個数を  $N_F$  個に拡大した後、帯域制限型の  $O_S$  倍(例えば8倍)のオーバーサンプリングを施すことにより  $O_S$  倍の個数の振幅デー



タを求め、この $O_S$  倍の個数 ( $m_M \times 1 + O_S$  個) の振幅データを直線補間してさらに多くの $N_M$ 個 (例えば2048個) に拡張し、この $N_M$  個のデータを間引いて上記一定個数 $M$  (例えば44個) のデータに変換している。

【0046】このデータ数変換部109からのデータ (上記一定個数 $M$ 個の振幅データ) がベクトル量子化部110に送られて、所定個数のデータ毎にまとめられてベクトルとされ、ベクトル量子化が施される。ベクトル量子化部110からの量子化出力データは、出力端子111を介して取り出される。また、上記高精度のピッチサーチ部106からの高精度 (ファイン) ピッチデータは、ピッチ符号化部115で符号化され、出力端子112を介して取り出される。さらに、上記有声音/無声音判別部107からの有声音/無声音 ( $V/U$ ) 判別データは、出力端子113を介して取り出される。これらの各出力端子111~113からのデータは、所定の伝送フォーマットの信号とされて伝送される。

【0047】なお、これらの各データは、上記 $N$ サンプル (例えば256サンプル) のブロック内のデータに対して処理を施すことにより得られるものであるが、ブロックは時間軸上を上記 $L$ サンプルのフレームを単位として前進することから、伝送するデータは上記フレーム単位で得られる。すなわち、上記フレーム周期でピッチデータ、 $V/U$ 判別データ、振幅データが更新されることになる。

【0048】次に、伝送されて得られた上記各データに基づき音声信号を合成するための合成側 (デコード側) の概略構成について、図15を参照しながら説明する。この図15において、入力端子121には上記ベクトル量子化された振幅データが、入力端子122には上記符号化されたピッチデータが、また入力端子123には上記 $V/U$ 判別データがそれぞれ供給される。入力端子121からの量子化振幅データは、逆ベクトル量子化部124に送られて逆量子化され、データ数逆変換部12

$$V_m(n) = A_m(n) \cos(\theta_m(n)) \quad 0 \leq n < L \quad \dots (11)$$

と表すことができる。全バンドの内の $V$  (有声音) と判別された全てのバンドの有声音を加算 ( $\sum V_m(n)$ ) して最終的な有声音 $V(n)$  を合成する。

【0051】この(11)式中の $A_m(n)$  は、上記合成フレームの先端から終端までの間で補間された第 $m$ 高調波の振幅である。最も簡単には、フレーム単位で更新され

$$A_m(n) = (L-n)A_{0m}/L + nA_{Lm}/L \quad \dots (12)$$

の式により $A_m(n)$  を計算すればよい。

$$\theta_m(0) = m\omega_{01}n + n^2 m(\omega_{L1} - \omega_{01}) / 2L + \phi_{0m} + \Delta\omega n \quad \dots (13)$$

により求めることができる。この(13)式中で、 $\phi_{0m}$  は上記合成フレームの先端 ( $n=0$ ) での第 $m$ 高調波の位相 (フレーム初期位相) を示し、 $\omega_{01}$  は合成フレーム先端 ( $n=0$ ) での基本角周波数、 $\omega_{L1}$  は該合成フレーム

5に送られて逆変換される。このデータ数逆変換部125では、上述した図1~図9の説明と対照的な逆変換が行われ、得られた振幅データが有声音合成部126及び無声音合成部127に送られる。入力端子122からの符号化ピッチデータは、ピッチ復号化部128で復号化され、データ数逆変換部125、有声音合成部126及び無声音合成部127に送られる。また入力端子123からの $V/U$ 判別データは有声音合成部126及び無声音合成部127に送られる。

【0049】有声音合成部126では例えば余弦( $\cos$ )波合成により時間軸上の有声音波形を合成し、無声音合成部127では例えばホワイトノイズをバンドパスフィルタでフィルタリングして時間軸上の無声音波形を合成し、これらの各有声音合成波形と無声音合成波形とを加算部129で加算合成して、出力端子130より取り出すようにしている。この場合、上記振幅データ、ピッチデータ及び $V/U$ 判別データは、上記分析時の1フレーム ( $L$ サンプル、例えば160サンプル) 毎に更新されて与えられるが、フレーム間の連続性を高める (円滑化する) ために、上記振幅データやピッチデータの各値を1フレーム中の例えば中心位置における各データ値とし、次のフレームの中心位置までの間 (合成時の1フレーム) の各データ値を補間により求める。すなわち、合成時の1フレーム (例えば上記分析フレームの中心から次の分析フレームの中心まで) において、先端サンプル点での各データ値と終端 (次の合成フレームの先端) サンプル点での各データ値とが与えられ、これらのサンプル点間の各データ値を補間により求めるようにしている。

【0050】以下、有声音合成部126における合成処理を詳細に説明する。上記 $V$  (有声音) と判別された第 $m$ バンド (第 $m$ 高調波の帯域) における時間軸上の上記1合成フレーム ( $L$ サンプル、例えば160サンプル) 分の有声音を $V_m(n)$  とするとき、この合成フレーム内の時間インデックス (サンプル番号)  $n$  を用いて、

る振幅データの第 $m$ 高調波の値を直線補間すればよい。すなわち、上記合成フレームの先端 ( $n=0$ ) での第 $m$ 高調波の振幅値を $A_{0m}$ 、該合成フレームの終端 ( $n=L$ : 次の合成フレームの先端) での第 $m$ 高調波の振幅値を $A_{Lm}$ とすると、

【0052】次に、上記(11)式中の位相 $\theta_m(n)$  は、

の終端 ( $n=L$ : 次の合成フレーム先端) での基本角周波数をそれぞれ示している。上記(13)式中の $\Delta\omega$  は、 $n=L$ における位相 $\phi_{Lm}$ が $\theta_m(L)$ に等しくなるような最小の $\Delta\omega$ を設定する。

【0053】以下、任意の第 $m$ バンドにおいて、それぞれ $n=0$ 、 $n=L$ のときの $V/U$ 判別結果に応じた上記振幅 $A_m(n)$ 、位相 $\theta_m(n)$ の求め方を説明する。第 $m$ バンドが、 $n=0$ 、 $n=L$ のいずれも $V$ （有声音）とされる場合に、振幅 $A_m(n)$ は、上述した(12)式により、伝送された振幅値 $A_{0m}$ 、 $A_{Lm}$ を直線補間して振幅 $A_m(n)$ を算出すればよい。位相 $\theta_m(n)$ は、 $n=0$ で $\theta_m(0)=\phi_{0m}$ から $n=L$ で $\theta_m(L)=\phi_{Lm}$ となるように $\Delta\omega$ を設定する。

【0054】次に、 $n=0$ のとき $V$ （有声音）で、 $n=L$ のとき $UV$ （無声音）とされる場合に、振幅 $A_m(n)$

$$\theta_m(0) = \phi_{Lm} - m(\omega_{01} + \omega_{L1})L/2 \quad \dots (14)$$

とし、かつ $\Delta\omega=0$ とする。

【0056】上記 $n=0$ 、 $n=L$ のいずれも $V$ （有声音）とされる場合に、 $\theta_m(L)$ が $\phi_{Lm}$ となるように $\Delta\omega$

$$\begin{aligned} \theta_m(L) &= m\omega_{01}L + L^2 m(\omega_{L1} - \omega_{01})/2L + \phi_{0m} + \Delta\omega L \\ &= m(\omega_{01} + \omega_{L1})L/2 + \phi_{0m} + \Delta\omega L \\ &= \phi_{Lm} \end{aligned}$$

となり、これを整理すると、 $\Delta\omega$ は、

$$\Delta\omega = (\text{mod } 2\pi((\phi_{Lm} - \phi_{0m}) - mL(\omega_{01} + \omega_{L1})/2))/L \quad \dots (15)$$

となる。この(15)式で $\text{mod } 2\pi(x)$ とは、 $x$ の主値を $-\pi \sim +\pi$ の間の値で返す関数である。例えば、 $x=1.3\pi$ のとき $\text{mod } 2\pi(x) = -0.7\pi$ 、 $x=2.3\pi$ のとき $\text{mod } 2\pi(x) = 0.3\pi$ 、 $x=-1.3\pi$ のとき $\text{mod } 2\pi(x) = 0.7\pi$ 、等である。

【0057】ここで、図16のAは、音声信号のスペクトルの一例を示しており、バンド番号（ハーモニクスナンバー） $m$ が8、9、10の各バンドが $UV$ （無声音）とされ、他のバンドは $V$ （有声音）とされている。この $V$ （有声音）のバンドの時間軸信号が上記有声音合成部126により合成され、 $UV$ （無声音）のバンドの時間軸信号が無声音合成部127で合成されるわけである。

【0058】以下、無声音合成部127における無声音合成処理を説明する。ホワイトノイズ発生部131からの時間軸上のホワイトノイズ信号波形を、所定の長さ（例えば256サンプル）で適当な窓関数（例えばハミング窓）により窓かけをし、STFT処理部132によりSTFT（ショートタイムフーリエ変換）処理を施すことにより、図16のBに示するようなホワイトノイズの周波数軸上のパワースペクトルを得る。このSTFT処理部132からのパワースペクトルをバンド振幅処理部133に送り、図16のCに示すように、上記 $UV$ （無声音）とされたバンド（例えば $m=8$ 、9、10）について上記振幅 $|A_m|_{UV}$ を乗算し、他の $V$ （有声音）とされたバンドの振幅を0にする。このバンド振幅処理部133には上記振幅データ、ピッチデータ、 $V/U$ 判別データが供給されている。バンド振幅処理部133からの出力は、ISTFT処理部134に送られ、位相は元のホワイトノイズの位相を用いて逆STFT処理を施

は、 $A_m(0)$ の伝送振幅値 $A_{0m}$ から $A_m(L)$ で0となるように直線補間する。 $n=L$ での伝送振幅値 $A_{Lm}$ は無声音の振幅値であり、後述する無声音合成の際に用いられる。位相 $\theta_m(n)$ は、 $\theta_m(0)=\phi_{0m}$ とし、かつ $\Delta\omega=0$ とする。

【0055】さらに、 $n=0$ のとき $UV$ （無声音）で、 $n=L$ のとき $V$ （有声音）とされる場合には、振幅 $A_m(n)$ は、 $n=0$ での振幅 $A_m(0)$ を0とし、 $n=L$ で伝送された振幅値 $A_{Lm}$ となるように直線補間する。位相 $\theta_m(n)$ については、 $n=0$ での位相 $\theta_m(0)$ として、フレーム終端での位相値 $\phi_{Lm}$ を用いて、

を設定する手法について説明する。上記(13)式で、 $n=L$ と置くことにより、

すことにより時間軸上の信号に変換する。ISTFT処理部134からの出力は、オーバーラップ加算部135に送られ、時間軸上で適当な（元の連続的なノイズ波形を復元できるように）重み付けをしながらオーバーラップ及び加算を繰り返し、連続的な時間軸波形を合成する。オーバーラップ加算部135からの出力信号が上記加算部129に送られる。

【0059】このように、各合成部126、127において合成されて時間軸上に戻された有声音部及び無声音部の各信号は、加算部129により適当な固定の混合比で加算して、出力端子130より再生された音声信号を取り出す。

【0060】なお、上記図10の音声分析側（エンコード側）の構成や図15の音声合成側（デコード側）の構成については、各部をハードウェア的に記載しているが、いわゆるDSP（デジタル信号プロセッサ）等を用いてソフトウェアプログラムにより実現することも可能である。

【0061】なお、本発明は上記実施例のみに限定されるものではなく、例えば、音声信号のみならず、音響信号を入力信号として用いることもできる。

【0062】

【発明の効果】以上の説明から明らかなように、本発明に係るデータ数変換方法によれば、ブロック内で可変個数のデータを非線形圧縮し、一定個数に変換しているため、ブロック間（フレーム間）の差分をとることやベクトル量子化等が可能となり、符号化効率を高める上で有効である。また、データ数変換（サンプル数変換）のための帯域制限型オーバーサンプリング処理を施す際に、

処理前のブロック内データの最後のデータ値と最初のデータ値との間を補間するようなダミーデータを付加してデータ個数を拡大しているため、後のフィルタ処理等により端点にリングングが発生するような不具合を回避でき、良好な符号化、特に効率の高いベクトル量子化が実現できる。

【図面の簡単な説明】

【図1】本発明に係るデータ数変換方法を説明するための概略構成を示すブロック図である。

【図2】データ数変化の一例を説明するための波形図である。

【図3】FFT前のデータ数拡大した波形の一例を示す図である。

【図4】FFT前のデータ数拡大した波形の比較例を示す図である。

【図5】FFT後の波形とオーバーサンプリング操作を説明するための図である。

【図6】FFT後の波形に対するフィルタリング操作を説明するための図である。

【図7】IFFT後の波形を示す図である。

【図8】オーバーサンプリングによるサンプル数変換の一例を示す図である。

【図9】直線補間及び間引き処理を説明するための図である。

【図10】本発明に係るデータ数変換方法が適用される装置の具体例としての音声信号の合成分析符号化装置の分析側（エンコード側）の概略構成を示す機能ブロック図である。

【図11】窓かけ処理を説明するための図である。

【図12】窓かけ処理と窓関数との関係を示すための図である。

【図13】直交変換（FFT）処理対象としての時間軸

データを示す図である。

【図14】周波数軸上のスペクトルデータ、スペクトル包絡線（エンベロープ）及び励起信号のパワースペクトルを示す図である。

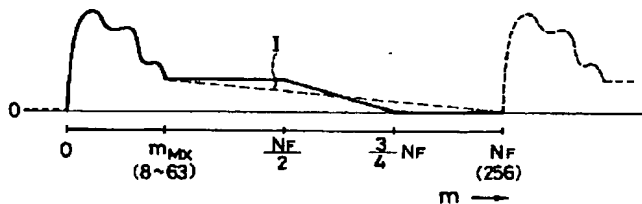
【図15】本発明に係るデータ数変換方法が適用される装置の具体例としての音声信号の合成分析符号化装置の合成側（デコード側）の概略構成を示す機能ブロック図である。

【図16】音声信号を合成する際の無声音合成を説明するための図である。

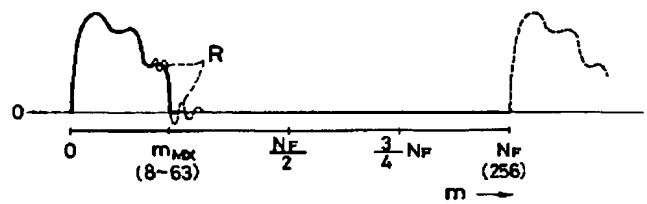
【符号の説明】

- 1 2 . . . . . 非線形圧縮部
- 1 3 . . . . . データ個数変換本体部
- 1 4 . . . . . ダミーデータ付加部
- 1 5 . . . . . 帯域制限型オーバーサンプリング部
- 1 6 . . . . . FFT（高速フーリエ変換）処理部
- 1 7 . . . . . 中間0詰め処理部
- 1 8 . . . . . IFFT（逆FFT）処理部
- 1 9 . . . . . 直線補間部
- 2 0 . . . . . 間引き処理部
- 1 0 3 . . . . . ピッチ抽出部
- 1 0 4 . . . . . 窓かけ処理部
- 1 0 5 . . . . . 直交変換（FFT）部
- 1 0 6 . . . . . 高精度（ファイン）ピッチサーチ部
- 1 0 7 . . . . . 有声音／無声音（V／UV）判別部
- 1 0 8 . . . . . 振幅再評価部
- 1 0 9 . . . . . データ数変換（データレートコンバー）部
- 1 1 0 . . . . . ベクトル量子化部
- 1 2 6 . . . . . 有声音合成部
- 1 2 7 . . . . . 無声音合成部

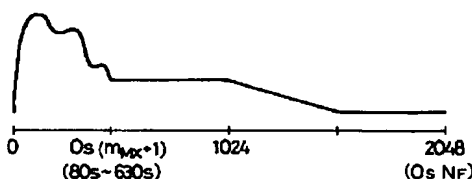
【図3】



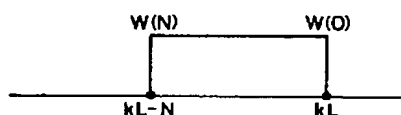
【図4】



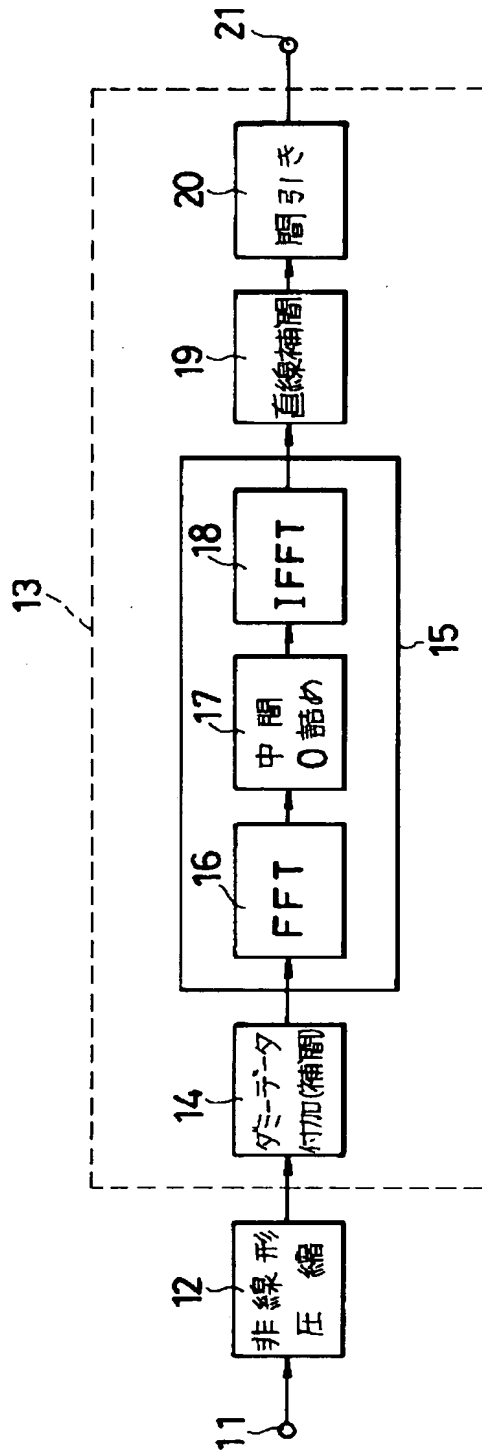
【図7】



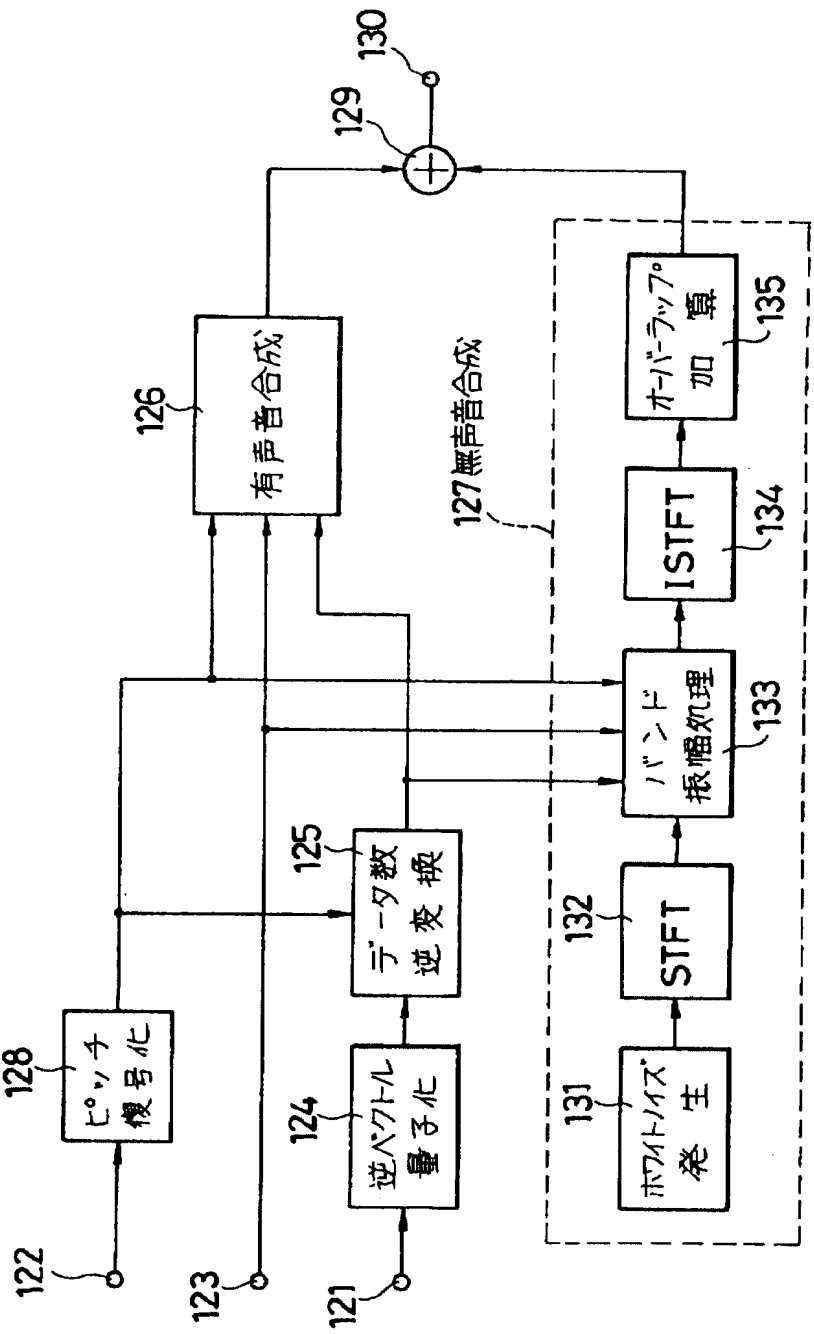
【図12】



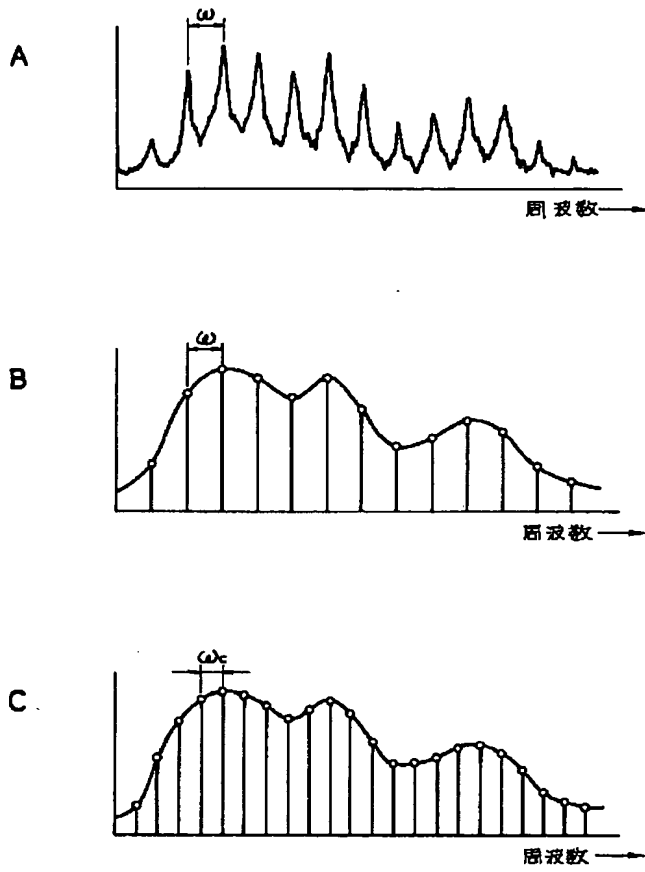
【図1】



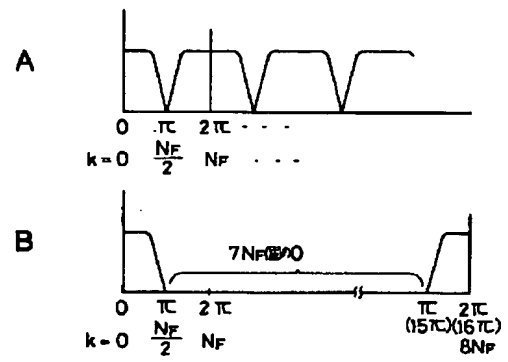
【図15】



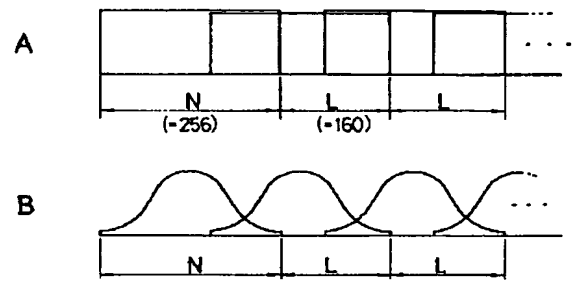
【図2】



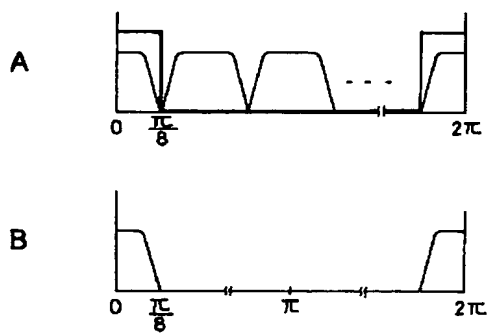
【図5】



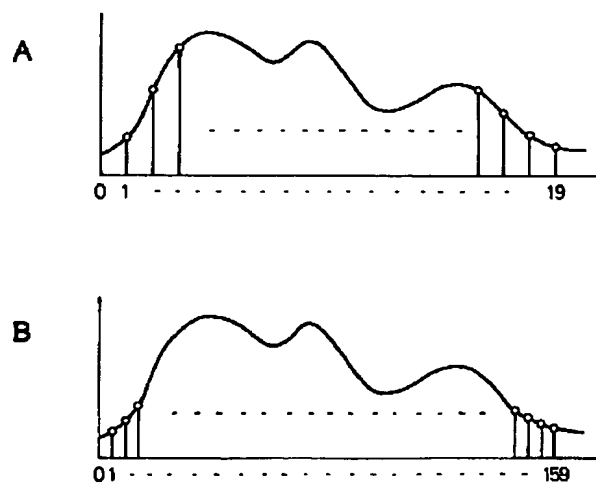
【図11】



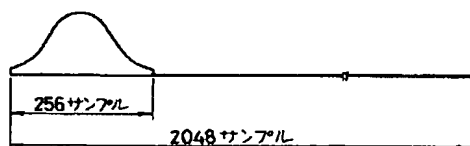
【図6】



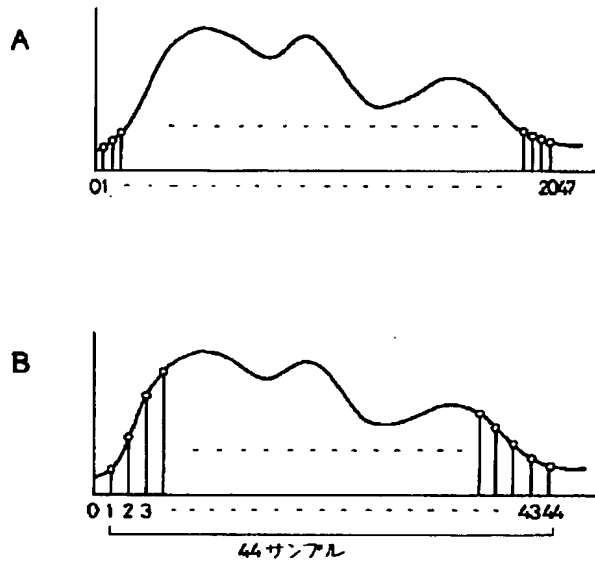
【図8】



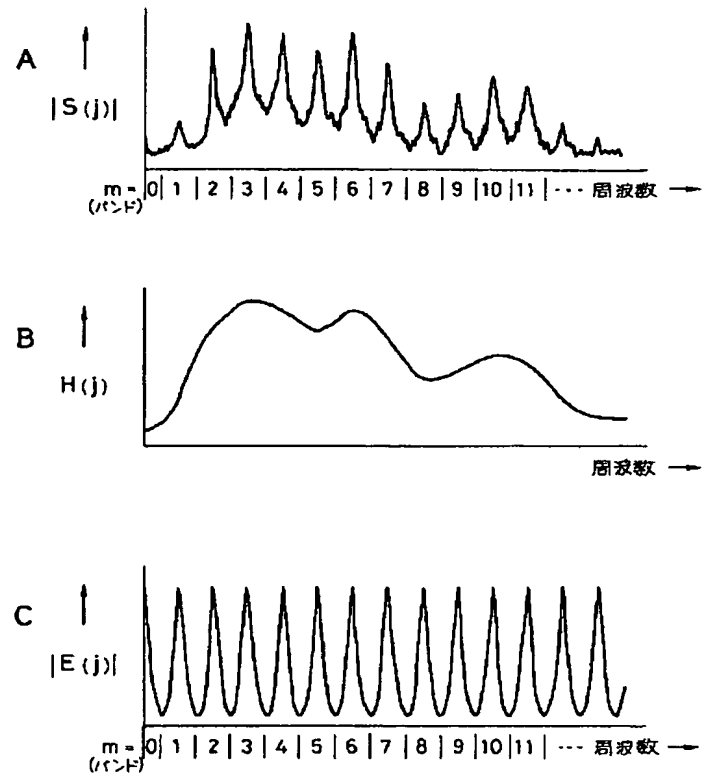
【図13】



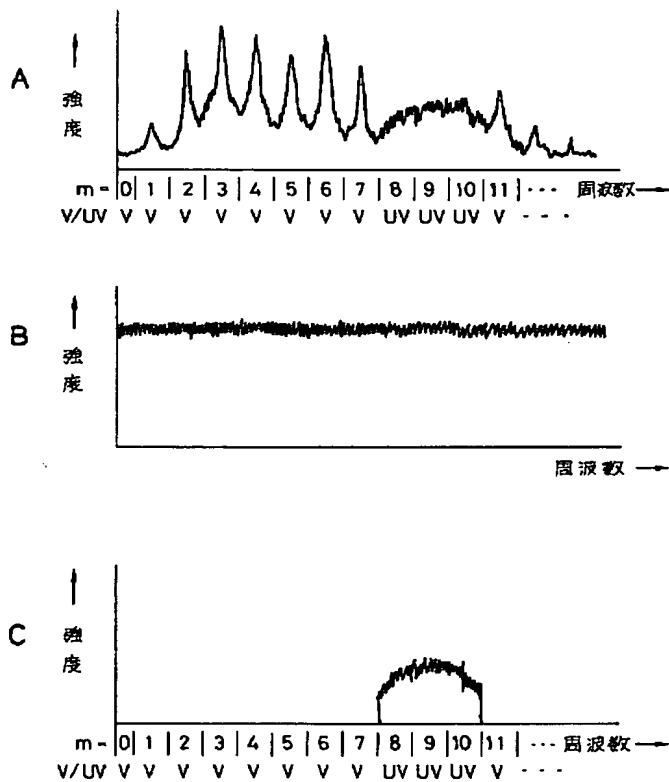
【図9】



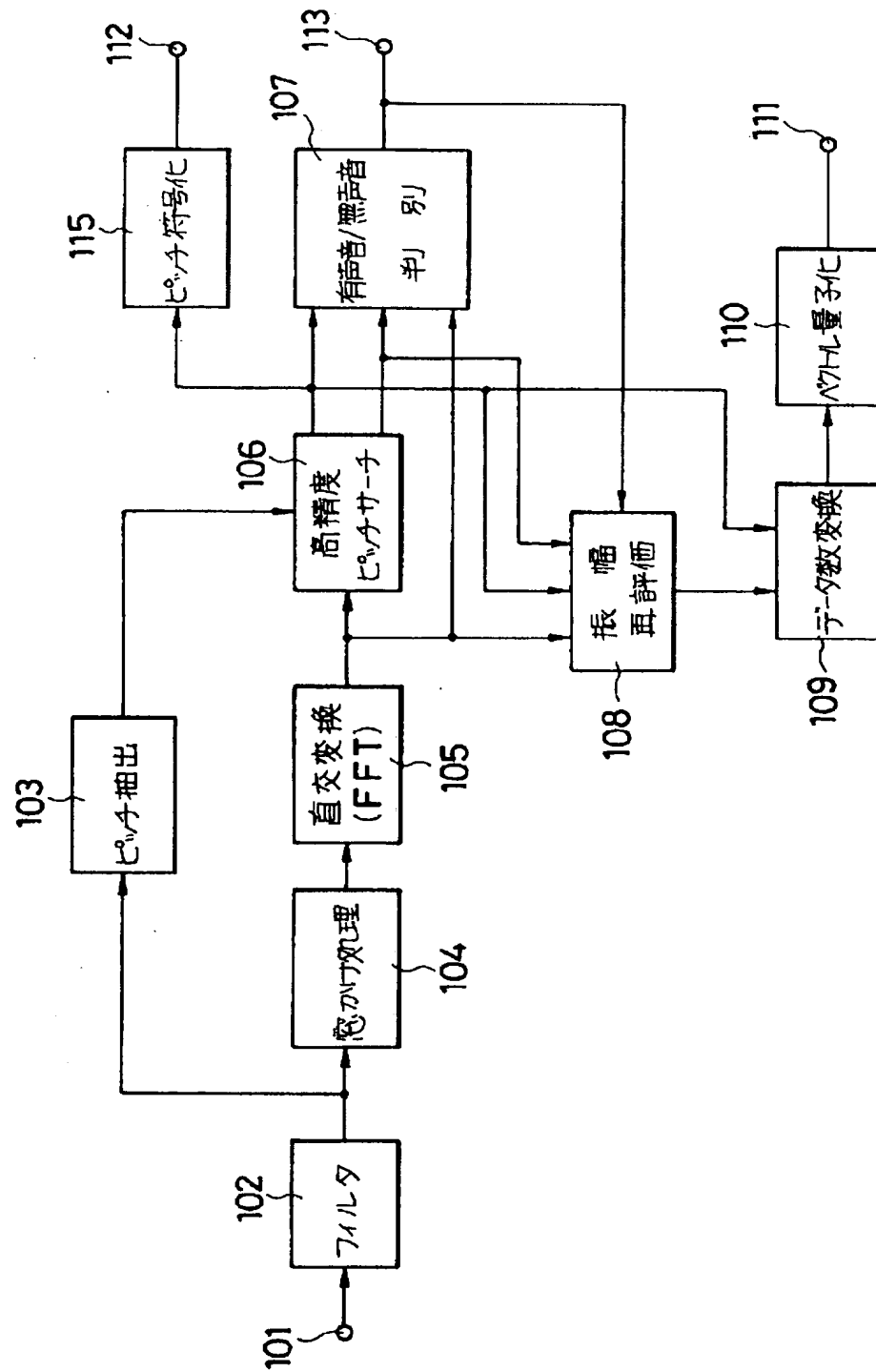
【図14】



【図16】



【図10】



## 【手続補正書】

【提出日】平成5年6月8日

## 【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0026

【補正方法】変更

【補正内容】

【0026】受信側（合成側あるいはデコーダ側）では、そのインデックスより、ベクトル量子化及び逆量子化された数列  $b_{VQdB}(n)$  なる  $M$  点の波形データを得る。そのデータ列を、同様の方法で、すなわち帯域制限オーバーサンプリング、直線補間、間引きの操作を施すことにより、必要なポイント数の上記  $m_{MX}+1$  点の数列に変換する。なお、 $m_{MX}$ （あるいは  $m_{MX}+1$ ）は、別途伝送するピッチ情報により求めることができる。例えばサンプリング周期に対して規格化されたピッチ周期を  $p$  とするとき、ピッチ周波数（角周波数） $\omega$  は、 $2\pi/p$  で求められ、 $\pi/\omega = p/2$  より、 $m_{MX}+1 = \text{int}(p/2)$  として算出できる。この  $m_{MX}+1$  点の振幅情報をもとにしてデコード処理を行う。

## 【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0031

【補正方法】変更

【補正内容】

【0031】ピッチ抽出部103では、上記  $X_{wr}(k, r)$  のサンプル列（1ブロック  $N$  サンプル）に基づいてピッチ抽出が行われる。このピッチ抽出法には、時間波形の周期性や、スペクトルの周期的周波数構造や、自己相関関数を用いるもの等が知られているが、本実施例では、センタクリップ波形の自己相関法を採用している。このときのブロック内でのセンタクリップレ

$$S(j) = H(j) \cdot |E(j)| \quad 0 \leq j < J \quad \dots (6)$$

と表現するようなモデルを想定している。ここで、 $J$  は  $\frac{\omega_s \angle 4\pi}{f_s} = f_s \cdot \angle 2$  に対応し、サンプリング周波数  $f_s = \frac{\omega_s \angle 2\pi}{2\pi}$  が例えば  $8 \text{ kHz}$  のときには  $4 \text{ kHz}$  に対応する。上記 (6) 式中において、周波数軸上のスペクトルデータ  $S(j)$  が図14のAに示すような波形のとき、 $H(j)$  は、図14のBに示すような元のスペクトルデータ  $S(j)$  のスペクトル包絡線（エンベロープ）を示し、 $E(j)$  は、図14のCに示すような等レベルで周期的な励起信号（エキサイテーション）のスペクトルを示している。すなわち、FFTスペクトル  $S(j)$  は、スペクトルエンベロープ  $H(j)$  と励起信号

$$\varepsilon_m = \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| \cdot |E(j)| \}^2 \quad \dots (7)$$

で表せる。このエラー  $\varepsilon_m$  を最小化するような  $|A_m|$  は、

ベルについては、1ブロックにつき1つのクリップレベルを設定してもよいが、ブロックを細分割した各部（各サブブロック）の信号のピークレベル等を検出し、これらの各サブブロックのピークレベル等の差が大きいときに、ブロック内でクリップレベルを段階的にあるいは連続的に変化させるようにしている。このセンタクリップ波形の自己相関データのピーク位置に基づいてピッチ周期を決めている。このとき、現在フレームに属する自己相関データ（自己相関は1ブロック  $N$  サンプルのデータを対象として求められる）から複数のピークを求めておき、これらの複数のピークの内の最大ピークが所定の閾値以上のときには該最大ピーク位置をピッチ周期とし、それ以外のときには、現在フレーム以外のフレーム、例えば前後のフレームで求められたピッチに対して所定の関係を満たすピッチ範囲内、例えば前フレームのピッチを中心として  $\pm 20\%$  の範囲内にあるピークを求め、このピーク位置に基づいて現在フレームのピッチを決定するようにしている。このピッチ抽出部103ではオープンループによる比較的ラフなピッチのサーチが行われ、抽出されたピッチデータは高精度（ファイン）ピッチサーチ部106に送られて、クローズドループによる高精度のピッチサーチ（ピッチのファインサーチ）が行われる。

## 【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0033

【補正方法】変更

【補正内容】

【0033】このピッチのファインサーチについて説明する。まず、上記MBEボコーダにおいては、上記FFT等により直交変換された周波数軸上のスペクトルデータとしての  $S(j)$  を

のパワースペクトル  $|E(j)|$  との積としてモデル化される。

## 【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0036

【補正方法】変更

【補正内容】

【0036】

【数1】

## 【手続補正5】

【補正対象書類名】明細書



【補正対象項目名】 0 0 3 7

【0 0 3 7】

【補正方法】 変更

【数 2】

【補正内容】

$$\frac{\partial \varepsilon_m}{\partial |A_m|} = -2 \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} |E(j)|$$

$$= 0$$

$$\therefore |A_m| = \frac{\sum_{j=a_m}^{b_m} |S(j)| |E(j)|}{\sum_{j=a_m}^{b_m} |E(j)|}^2$$

... (8)

となり、この(8)式の $|A_m|$ のとき、エラー $\varepsilon_m$ を最小化する。このような振幅 $|A_m|$ を各バンド毎に求め、得られた各振幅 $|A_m|$ を用いて上記(7)式で定義された各バンド毎のエラー $\varepsilon_m$ を求める。次に、このような各バンド毎のエラー $\varepsilon_m$ の全バンドの総和値 $\sum \varepsilon_m$ を求める。さらに、このような全バンドのエラー総和値 $\sum \varepsilon_m$ を、いくつかの微小に異なるピッチについて求め、エラー総和値 $\sum \varepsilon_m$ が最小となるようなピッチを求める。

【手続補正6】

【補正対象書類名】 明細書

【補正対象項目名】 0 0 4 8

【補正方法】 変更

【補正内容】

【0 0 4 8】次に、伝送されて得られた上記各データに基づき音声信号を合成するための合成側（デコード側）の概略構成について、図15を参照しながら説明する。この図15において、入力端子121には上記ベクトル量子化された振幅データが、入力端子122には上記符号化されたピッチデータが、また入力端子123には上

$$\theta_m(n) = m\omega_{01}n + n^2m(\omega_{L1} - \omega_{01}) / 2L + \phi_{0m} + \Delta\omega n$$

$$\dots (13)$$

により求めることができる。この(13)式で、 $\phi_{0m}$ は上記合成フレームの先端( $n=0$ )での第 $m$ 高調波の位相（フレーム初期位相）を示し、 $\omega_{01}$ は合成フレーム先端( $n=0$ )での基本角周波数、 $\omega_{L1}$ は該合

記V/U判別データがそれぞれ供給される。入力端子121からの量子化振幅データは、逆ベクトル量子化部124に送られて逆量子化され、データ数逆変換部125に送られて逆変換される。このデータ数逆変換部125では、上述した図1～図9の説明と同様な変換が行われ、得られた振幅データが有声音合成部126及び無声音合成部127に送られる。入力端子122からの符号化ピッチデータは、ピッチ復号化部128で復号化され、データ数逆変換部125、有声音合成部126及び無声音合成部127に送られる。また入力端子123からのV/U判別データは有声音合成部126及び無声音合成部127に送られる。

【手続補正7】

【補正対象書類名】 明細書

【補正対象項目名】 0 0 5 2

【補正方法】 変更

【補正内容】

【0 0 5 2】次に、上記(11)式中の位相 $\theta_m(n)$ は、

成フレームの終端( $n=L$ :次の合成フレーム先端)での基本角周波数をそれぞれ示している。上記(13)式中の $\Delta\omega$ は、 $n=L$ における位相 $\phi_{Lm}$ が $\theta_m(L)$ に等しくなるような最小の $\Delta\omega$ を設定する。